

INTRODUCTION

Corpus Linguistics has been slowly breaking ground in Brazil's academic scene, especially in the field of translation, where corpora were until recently as good as unknown. As a contribution to divulge this area, I edited in 2003 "Translation and Corpora", a special issue of *Cadernos de Tradução*, published by the Federal University of Santa Catarina, Brazil.

This issue of TradTerm comes out in the wake of "Translation and Corpora" and aims to present a picture of how corpora are being used in translation-related research, especially – but not only – in Brazil. The articles included in this issue focus either on existing corpora or on corpora in the process of being built, discussing the uses they can be put to or reporting on studies they have been used for.

Sara Laviosa, from the University of Bari, Italy, compiler of the ECC – English Comparable Corpus (1996), offers a detailed panorama of the translation-corpora partnership we are addressing in this issue. Starting with Baker's seminal article (1993), she outlines the evolution of Corpus-Based Translation Studies (CTS) and discusses the period from 1996 to 1999, which came to be known as "the corpus linguistic turn in Translation Studies" in contrast with the period from 2000 onwards, when the author believes "it is plausible to envisage a cultural studies turn in CTS." The article examines the relationships between Descriptive Translation Studies (DTS), CTS, and Corpus Linguistics so as to assess which proposals from the past are still valid and which areas of investigation are most promising within CTS nowadays.

Some of the articles discuss the construction of various corpora, such as Stig Johansson's, from the Department of British and American Studies, University of Oslo, Norway. Johansson is one of the authors of the LOB Corpus (Lancaster-Oslo/Bergen Corpus), compiled in 1978, and one of the first European

¹ In this case a parallel corpus is understood as a corpus of original texts and their corresponding translations

researchers to build a bi-directional parallel corpus, the ENPC (English-Norwegian Parallel Corpus), in 1994.¹ His article discusses the model underlying the corpus and the methodology followed in building it, and offers a detailed account of the different ways in which the corpus can be queried with the Translation Corpus Explorer, a program especially developed to be used with it. Finally, Johansson discusses how a language can be studied by means of translation and how translation can be studied with the help of corpora.

Contrastive Studies, one of the areas of investigation singled out by Johansson, is explored by Josef Schmied, director of the REAL Centre (Research in English and Applied Linguistics) at the Chemnitz University of Technology, Germany. Schmied is particularly interested in the use of natural language in various contexts: historical, social, regional, and stylistic. His studies are based on the Chemnitz English-German Translation Corpus, likewise a bi-directional parallel corpus. Schmied's article also presents the Chemnitz Internet Grammar, a learning tool that allows users to investigate English grammar both inductively and deductively. This tool was created in 1999 and upgraded in 2003; its analyses are based on the corpus mentioned above. Schmied stresses the use of translation corpora as a discovery procedure especially suitable for non-native speakers, since its results are based on a large amount of data and therefore do not depend on the researcher's introspection. Schmied also points out that bi-directional parallel corpora are of great value for contrastive studies, as they bring out linguistic preferences in one language and enable studies of phenomena involving gradation, like the auxiliary-catenative-full verb cline. These possibilities are illustrated from a contrastive point of view with case studies of auxiliary *help*, catenatives *appear/seem* and their German equivalents, and the deontic-epistemic spectrum of modal *may/might* with their corresponding German structures. Despite acknowledging the limitations that corpora may present, Schmied highlights their role in contrastive studies, arguing that keywords in context may disclose "more grammar than meets the eye".

Stella E. O. Tagnin, a translation professor and researcher at the Department of Modern Languages, University of São Paulo,

Brazil, has published, among others, in the areas of Corpus Studies, English phraseology and literary translation.² In the present article she introduces the COMET project – A Multilingual Corpus for Teaching and Translation, which she coordinates. COMET consists of three subcorpora: the CorTeC (Technical-Scientific Corpus), a Translation Corpus and a Learner Corpus.

CorTeC is a bilingual English-Portuguese corpus consisting of subcorpora in several technical and scientific fields with special emphasis on four of them: Computing, Orthodontics, Environment and Commercial Law. The subcorpora are comparable corpora, therefore consisting of original texts in both languages, which will permit the investigation of the phraseology and terminology peculiar to those fields. The Translation Corpus is subdivided into a parallel corpus and a comparable corpus. The former includes literary and technical-scientific texts and contains originals and their corresponding translations, which will enable both research into translation strategies and norms, and contrastive lexico-grammatical studies. The comparable corpus is composed of translations produced by sworn translators in several languages. In Brazil the registry books with every sworn translator's lifetime output is deposited with the corresponding *Junta Comercial*³ after his retirement. This corpus is described in detail in the article by Aubert & Tagnin (see below). The Learner Corpus, consisting of texts written by foreign language learners, will not be dealt with here as it bears no relation to translation.

Another corpus being built in Brazil is CORDIALL, which is introduced in the article by Adriana Pagano, Célia Maria Magalhães and Fábio Alves from the Federal University of Minas Gerais, Brazil. Adriana Pagano coordinates two projects at present: "Corpora, cognition and discourse: an interdisciplinary proposal for

² Tagnin, S.E.O. *Expressões idiomáticas e convencionais*. São Paulo, Ática, 1989.

_____. *Contos Canadenses*. São Paulo, Olavobrás, 2002.

_____. (ed.) "Tradução e Corpora" número especial de *Cadernos de Tradução*. Florianópolis, Universidade Federal de Santa Catarina, vol. IX/2002/1.

³ The official commercial office in charge, among others, of the accreditation of sworn translators.

translation studies based on electronic databanks” and “Corpora, gender and (re)textualization: interfaces in translation studies”. Célia Magalhães is a professor of English and translation. She is mainly interested in investigating cohesive links and their discursive role in parallel English-Portuguese corpora. She is the author of *Os Monstros e a Questão Racial na Literatura Modernista Brasileira*⁴ (“Monsters and the Racial Issue in Brazilian Modernist Literature”) and the editor of *Reflexões sobre a Análise Crítica do Discurso*⁵ (“Essays on Critical Analysis of Discourse”). She has published in the journals *Cadernos de Tradução*, *DELTA*, *CROP*, and *TradTerm*, and contributed with chapters of books published and translated in Brazil. Fabio Alves is a professor of German and translation. He is mainly interested in studying the translation process and the cognitive aspects involved in the acquisition of translation competence. He is the author of a large number of publications in Brazil and abroad.⁶ Together, the three authors published a book on strategies for translation trainees to develop their autonomy.⁷

CORDIALL consists of four subcorpora. The first one is a parallel corpus with originals and their corresponding translations for the following language pairs: Portuguese-Spanish, Portuguese-German, Portuguese-English, and Spanish-English. It is aimed at carrying out comparative analyses of translators’ decisions and strategies in their re-textualizations as well as studies of the historical parameters involved in translation production. The second one is a comparable corpus of Brazilian Portuguese with original and translated texts, a particularly helpful setup for the study of the peculiarities of translated language.

⁴ Belo Horizonte, Editora UFMG, 2003.

⁵ Belo Horizonte, POSLIN/FALE/UFMG, 2001.

⁶ Among others:

Alves, F. *Zwischen Schweigen und Sprechen: Wie bildet sich eine transkulturelle Brücke?* Hamburg, Dr. Kovac, 1995.

Alves, F. (ed.) *Triangulating translation: perspectives in process oriented research*. Amsterdam, John Benjamins, 2003.

⁷ Alves, F.; Magalhães, C. & Pagano, A. *Traduzir com autonomia: estratégias para o tradutor em formação*. São Paulo, Contexto, 2000.

The third corpus, CORPRAT, is a multilingual processual corpus compiled so as to provide data for the investigation of inference patterns, strategic planning, solution of problems, and decision-making during the translation process. The fourth corpus is a specialized corpus with journalistic and academic texts. It is not specifically related to translation studies as it was built for rhetorical and lexical studies. In their article, the authors discuss the theoretical foundations on which these corpora have been built “as a resource for the study of discourse and cognitive issues in an interdisciplinary approach, drawing on insights from corpora studies, translation studies, cognitive studies, discourse analysis, and cultural studies.” CORDIALL is being constantly added to and restructured and aims at informing studies based on smaller corpora from a discursive and cognitive perspective.

A quite peculiar corpus is presented by Francis Henrik Aubert and Stella. E. O. Tagnin, both professors at the Department of Modern Languages, University of São Paulo, Brazil. Aubert is a sworn translator and has published extensively in the fields of translation and terminology. Tagnin, as mentioned before, is the coordinator of the COMET project, which hosts the corpus discussed in this article. The corpus consists of sworn translations produced over a one-hundred-year period (1902-2002), duly recorded in official books, which are sometimes in extremely bad conditions due to their age. Most of these translations are into Portuguese and cover a wide variety of text types, from personal documents, like birth certificates and transcripts of studies, to corporate by-laws and contracts of international cooperation. Because of the large number of technical problems involved in digitalizing the material, the construction of the corpus will begin with the most recent translations (1972-2002) and will restrict itself to the five languages taught at their department (English, French, German, Italian, and Spanish), out of the over twenty languages represented in the corpus. Detailed information in the header of each text will allow researchers to build their own subcorpus (by selecting language, textual type, domain, period, translator, etc.). The corpus will also enable interdisciplinary studies in History, Anthropology, Economics, and Politics.

TRADTERM, **10**, 2004, p. 19-28

Complementing the previous articles, concerned with the setup of corpora, the articles introduced below will present practical examples of how such reference materials can be used and explored.

The first paper takes a cognitive perspective showing a clear link with pedagogical issues. Fabio Alves and Célia Maria Magalhães, from the Federal University of Minas Gerais, Brazil, detect the need to map the process-product interface in translation and propose to investigate translator performance by means of triangulating data-eliciting techniques, like TAPs and Translog, which register all keyboard activity of a translator during the translation process, as well as WordSmith Tools, as applied to small corpora. The corpus used is CORPRAT (described in the article by Pagano, Magalhães and Alves above), which makes possible the study of cognitive and discursive aspects of translator performance. The aim of the research is to distinguish between performance features peculiar to novice translators in comparison with professional translators, so as to collect data that may inform a pedagogical reflection on translator training. The following are some of the topics of interest: translators' orientation, drafting, and revision patterns, their problem-solving and decision-making strategies, their "cognitive rhythm", as well as cognitive and discursive features related to a critical language awareness. The research involved the translation of a 63-word text from English into Portuguese by 17 subjects using Translog. The authors hypothesize that the orientation-drafting-revision phases of novice translators would vary widely and their cognitive rhythm would present an erratic pattern. A close analysis of the data reveals that there is no correlation between the novice translators' erratic rhythm and the quality of the texts they produce. A low level of language awareness and production of less durable texts were also observed. Furthermore, strict linear processing and lack of adequate cognitive administration and critical linguistic awareness seems to prevent novice translators from improving their work and producing more stable texts. Based on these results, the authors suggest ways of improving the training of novice translators so that their performance can come closer to the performance of professional translators.

Despite the increase in the use of corpora in translator training, the same is not true among professional translators. To investigate the reasons for this discrepancy Lynne Bowker, a professor at the Translation and Interpretation School of the University of Ottawa and the author of two books on the use of corpora in translator training,⁸ made a survey of publications by translation teachers and by Canadian translators' associations. The latter survey is complemented by a database of job advertisements aimed at analyzing the number of companies looking for translators with some corpus-related abilities. The discussion of the academic output considers both the research carried out with the use of corpora and the uses corpora are put to in the training of translators (e.g. translation memories). The material published by translators' associations refers mainly to events related to translation memories; conventional corpus tools have been the object of just one single event in Canada. However, hybrid tools, which combine a translation memory with corpus analysis tools, are coming on the market. From the job advertisement database, which includes a few other jobs like revisers, localization specialists and terminologists, Bowker extracts the abilities required from candidates, establishing a difference between general computational knowledge and acquaintance with translation memories and similar tools. The author concludes that the introduction of hybrid tools may enhance the use of corpora by professional translators but that, because of workplace constraints, especially in terms of practice and productivity, corpora will never be as popular among professionals as they are in academic circles, where emphasis is placed on research and transfer of knowledge. Nevertheless, her conclusions stress the fact that the knowledge acquired by corpus-based translator training can be easily transferred to the professional workplace.

The following four articles highlight research possibilities with bilingual corpora. They report on specific studies in different

⁸ Bowker, L. *Computer-aided translation technology: a practical Introduction*. Ottawa, University of Ottawa Press, 2002.

Bowker, L. & Pearson, J. *Working with specialized language: a practical guide to using corpora*. London/NY, Routledge, 2002.

language pairs, focusing on different linguistic levels and resorting to a corpus-based or a corpus-driven approach.

Karin Aijmer, from the English Department, University of Gothenburg, Sweden, is the co-author, with Bengt Altenberg, of the ESPC (English-Swedish Parallel Corpus) and of *English Corpus Linguistics* (1991), and one of the most productive researchers in Corpus Linguistics. Her article presents a contrastive semantic study of the polysemous verb “to see” and its equivalents in Swedish. Since the verbs are cognates in both languages and are both of Germanic origin, one would expect a high rate of instances in which the verbs are translated by their cognates. However, Aijmer found that in the English-Swedish direction there was a 69.3% correspondence, while in the opposite direction the correspondence was only 50.7%. A careful analysis showed that the various meanings of “to see” could be associated with different grammatical contexts. For instance, *see* followed by a clause introduced by *that* or *wh-* has a metalinguistic status and “refers to seeing as the source of knowledge for drawing a conclusion rather than to seeing itself.” With verbs of emotion, on the other hand (e.g. *I regret to see*), it has a factive meaning. An interesting feature Aijmer detects is the so-called process of “pragmaticalization”, as in *I see, you see*, where meaning is accounted for on the interpersonal or textual level. The author claims that the meaning extensions of *see* are culture-dependent, so that analysis of other languages may reveal new semantic extensions of this verb.

Sara Laviosa has contributed with a second article: a study of the anglicism *business* in Italian. The purpose of the study is to identify the contexts in which the English term may be translated by the anglicism in Italian. Laviosa uses two corpora for her investigation: COMIC (COMmercial Italian Corpus) and SALCA, both consisting of business texts, the first in Italian, the second in English. Initially, the author identifies four colligational patterns for *business* in COMIC and relates them to four distinct meanings. Next, she investigates SALCA, to see how these meanings are realized in English. Lastly, she suggests the contexts in which *business* may be translated by the loan-word *business* in Italian. The investigation is part of a broader study which will

deal with all synonyms of *business* in Italian, such as *affari*, *commercio*, *trattativa*, *transazione*, etc., in an attempt to suggest equivalents for this highly recurring term in Italian.

Another article which contrasts Italian and English, though from a sociocultural perspective, was written by Elena Tognini Bonelli and Elena Manca, both from the University of Lecce, Italy. Tognini Bonelli teaches at the Department of Philology, Linguistics and Literature, and is the author of *Linguistics at Work* (2001), a book which, due to its clear text and wealth of examples, should be compulsory reading for students carrying out research with corpora. The authors use comparable corpora addressing the language of tourism – *Farmhols*, in English, and *Agriturist*, in Italian – and describe a quite innovative methodology to arrive at a translation equivalent for *welcome*.⁹ Based on the concordances and collocates for *welcome* (*children*, *pets*, *dogs* and *guests*), the authors investigate the contexts of occurrence of their *prima facie* equivalents in Italian: *bambini*, *animali* and *ospiti*. Comparison of the syntagmatic patterns and paradigmatic associations of the translation equivalents allowed them to identify the functional equivalents, which, in turn, revealed clear cultural and ideological implications. The study also shows that, although single words may have a conventional translation in another language, this is not always the case when such words combine to form a lexical unit. In such cases, a larger stretch of context will suggest functionally equivalent translations.

The next article, by Stella E. O. Tagnin (introduced above) and Elisa Duarte Teixeira, a doctoral student in Translation at the University of São Paulo, Brazil presents empirical data to justify the affirmative answer they usually give to a frequently asked question: “Is Brazilian Portuguese (PB) very different from European Portuguese (PE)?” In fact, the authors claim that the difference is much more significant than that between American English (IA) and British English (IB). For that purpose they built a quadricultural comparable corpus of cooking recipes (Brazilian

⁹ *Welcome* is a high-frequency word in the English corpus (324 occurrences) but seems to have no cognate correspondent in Italian as *benvenuto*, its *prima facie* equivalent, which occurs only 4 times in the Italian corpus.

Portuguese [PB], European Portuguese [PE], American English [IA], and British English [IB]). Keyword analysis of each corpus and of each of their subdivisions (ingredients, preparation, comments, etc) and of their collocates and clusters have revealed significant differences at the lexical, syntactic and discursive levels. At the lexical level, while the analysis of the Portuguese and Brazilian texts shows 20 PE words with no occurrence in the PB corpus and 7 PB words with no occurrence in the PE corpus, the difference between the English variants is restricted to 6 words related to measurements (kg, l, ml, etc.) and only 4 content words (IB *cornflour*, *prawns*, *courgettes*, which do not occur in IA; and IA *ham*, which does not occur in IB). At the syntactic level, there are clear differences between PB and PE: the impersonal pronominal structure *deixa-se*, *junta-se* in PE, as opposed to the imperative *deixe*, *junte* in PB. At the discursive level, the data show that PB recipes usually have two distinct sections: “*ingredientes*” (ingredients) and “*modo de fazer*” (preparation), a structure which is not so common in PE. By observing certain similarities between Brazilian Portuguese and American English on the one hand, and between European Portuguese and British English on the other, the authors have advanced some cultural hypotheses, which will be tested on an enlarged version of the present corpus.

We hope that the selection of papers in this issue will contribute to make corpus studies related to Translation better known in Brazil, and will motivate translation researchers as yet unfamiliar with this methodology to embark on this promising and revealing interface.

São Paulo, March 2004
Stella E. O. Tagnin